

Jerzy Marzec

BAYESOWSKI MODEL DWUMIANOWY OPARTY NA MIESZANCE ROZKŁADÓW NORMALNYCH

1. Wprowadzenie

Spośród ekonometrycznych modeli danych jakościowych najbardziej znanymi są model logitowy i probitowy. W literaturze proponuje się specyfikacje z innymi typami rozkładów (zob. Maddala [1983]), ale w badaniach empirycznych te pierwsze są powszechnie stosowane. Modele logitowy i probitowy, zarówno w przypadku dychotomicznej zmiennej endogenicznej, jak i zmiennej wielomianowej, są bardzo dobrze opisane w literaturze, ich własności są zbadane, a estymacja metodą największej wiarygodności (MNW) jest dostępna w wielu programach statystycznych. Geweke i Keane [1999] zwracają uwagę, że podstawowym modelem danych jakościowych jest model probitowy, gdyż w przypadku kategorii nieuporządkowanych umożliwia uwzględnienie korelacji między parami alternatyw (wyborów). Natomiast wielomianowy model logitowy został poddany krytyce przez McFaddena [1984], gdyż zakłada niezależność ilorazu szans dwóch alternatyw od pozostałych wyborów (ang. *independence from irrelevant alternatives*) zob. także Amemiya [1985], Maddala [1983].

Jednym z możliwych kierunków uogólnienia modeli danych jakościowych jest stosowanie szerszej klasy rozkładów dla składnika losowego, np. w formie mieszanek rozkładów. Błędy specyfikacji na etapie założeń odnośnie struktury stochastycznej prowadzą do niezgodnych estymatorów parametrów strukturalnych. Wykorzystywanie modelu probitowego czy logitowego, gdy prawdziwy rozkład jest inny niż przyjęto, może prowadzić do błędnego wnioskowania m.in. o prawdopodobieństwach zaobserwowania poszczególnych wyborów i efektach krańcowych.

Modele statystyczne oparte na mieszanekach rozkładów statystycznych znane są od końca XIX wieku, dzięki pracy Karla Pearsona [1894]. Najczęściej stosuje się skończone mieszanki rozkładów normalnych. Ich zastosowanie, jako aproksymacji nieznanego postaci rozkładu dla składnika losowego, pozwala uwzględnić takie własności jak: wielomodalność, asymetrię oraz zachowanie się rozkładu w otoczeniu wartości modalnej (wyostrzenie albo spłaszczenie w odniesieniu do rozkładu normalnego). Stąd cieszą się one dużym zainteresowaniem. Roeder i Wassermann [1997] prezentują 10 możliwych typów kształtów rozkładów, które można uzyskać z tych mieszanek. Diebolt i Robert [1994] uważają, że te modele stanowią interesującą alternatywę dla modeli nieparametrycznych, gdyż w ramach swej klasy przyjmują słabe założenia dotyczące składnika losowego. Zatem mieszanki dają szeroki wachlarz możliwości, gdy klasyczne założenia nie są spełnione, np. w sytuacji niejednorodnej próby, występowania heteroscedastyczności itp. Mają one szczególne znaczenie wówczas, gdy istotny jest dobór odpowiedniej postaci rozkładu dla składnika losowego.

Z drugiej strony, choć mieszanki są atrakcyjne z punktu widzenia własności, to ich parametryczna analiza statystyczna nie jest łatwa, zwłaszcza od strony obliczeniowej. Problemy wynikają ze skomplikowanej postaci funkcji wiarygodności i konieczności uwzględnienia restrykcji na parametry zapewniających identyfikowalność. Aitkin [2001] zwraca uwagę, że przy specyficznej konstrukcji danych funkcja wiarygodności mieszanki może być nawet nieograniczona. Ten problem (*infinite spike*) może mieć

miejsce szczególnie, gdy liczba obserwacji jest niewiele większa od liczby składników mieszanki. Wówczas istnieje niezerowe prawdopodobieństwo, że próba nie niesie informacji o jednym ze składników mieszanki; zob. także Marin, Mengersen i Robert [2005].

W celu parametrycznej estymacji modeli mieszanek proponuje się trzy podstawowe metody: MNW, algorytm EM (*Expectation Maximization*) (Dempster, Laird i Rubin 1977) oraz Monte Carlo oparte na łańcuchach Markowa (MCMC – *Monte Carlo Markov Chain*). Ta ostatnia odgrywa istotną rolę we wnioskowaniu bayesowskim, gdyż jest skutecznym narzędziem numerycznego całkowania, zob. Tierney [1994]. Bayesowska estymacja modeli regresji z mieszanką jest omawiana m.in. w pracach: Diebolt i Robert [1994], Richardson i Green [1997], Roeder i Wassermann [1997], Celeux, Hurn i Robert [2000], Aitkin [2001], Frühwirth-Schnatter [2001], Jasra, Holmes i Stephens [2005], Marin, Mengersen i Robert [2005], Geweke [2007]. Wielu z tych autorów podkreśla, że powtarne i szerokie zainteresowanie modelami mieszanek, znanymi przecież od wielu dziesiątek lat, jest spowodowane rozwojem i upowszechnianiem się metod MCMC, które (dzięki rozwojowi informatyki) czynią wnioskowanie bayesowskie użytecznym i elastycznym podejściem w przypadku nawet bardzo skomplikowanych modeli.

W naukach przyrodniczych i inżynierskich mieszanki mają szerokie zastosowanie, a obszerną listę aplikacji można znaleźć w monografii Titterton, Schmidt i Makov [1985]. W ekonomii wykorzystuje się je w takich dziedzinach jak np.: marketing (Jedidi [1997], Allenby, Leone i Jen [1999]), makroekonomia (Hamilton [1989], Lesage [1992], So, Lan i Li [1998], Frühwirth-Schnatter [2001]), finanse (Dueker [1999], Liesefeld [2001]). Bardzo często znajdują one zastosowanie w badaniach empirycznych, które opierają się na ekonometrycznych modelach regresji przełącznikowej; zob. Quandt i Ramsey [1978] oraz wykaz literatury w pracy Geweke [2007]. Natomiast niewiele jest opracowań, w których rozważano by modele zmiennych jakościowych wykorzystując skończone mieszanki rozkładów normalnych. Erkanli, Stang i Müller [1993] oraz Frühwirth-Schnatter i Frühwirth [2007] wykorzystali mieszanki w analizie modeli wielomianowych, Geweke i Keane [1999] oraz Qu i Qu [2000] w modelu dwumianowym, zaś Austin i Escobar [2003] w modelu tobitowym. Na uwagę zasługuje fakt, że w tych pracach stosuje się wyłącznie wnioskowanie bayesowskie, które umożliwia pełną analizę statystyczną modeli z nieobserwowalnymi zmiennymi ukrytymi.

Celem niniejszego artykułu jest prezentacja modelu dwumianowego (dla binarnej zmiennej endogenicznej) opartego na skończonej mieszance rozkładów normalnych. Rozważamy przypadek mieszanek dwuskładnikowych, których estymacja zostanie zrealizowana w ramach podejścia bayesowskiego. To podejście jest szczególnie uzasadnione zwłaszcza w sytuacji, gdy badacz posiada silną wiedzę a priori o parametrach. W tym przypadku wynika ona z konieczności narzucenia restrykcji zapewniających identyfikowalność parametrów rozważanych modeli. Ponadto na gruncie bayesowskim możliwa jest pełna prezentacja i porównanie niepewności dotyczącej badanej wielkości, zarówno przed zaobserwowaniem zjawiska – w formie rozkładu a priori, jak po wglądzie w dane – poprzez rozkład a posteriori. Zatem omówimy dobór rozkładów a priori, który jest ważnym i trudnym zagadnieniem w tej klasie modeli. W literaturze, w przypadku modeli regresji opartych na mieszankach, jako numeryczną metodę uzyskiwania rozkładów a posteriori proponuje się przede wszystkim schemat Gibbsa; zob. Geweke [2007]. Natomiast Celeux, Hurn i Robert [2000] poddają krytyce jego stosowanie. W przypadku modelu dwumianowego, gdy próba jest duża, stosowanie próbnika Gibbsa jest czasochłonne. Z tego też powodu wykorzystaliśmy algorytm Metropolisa i Hastingsa.

Wprowadzone uogólnienie w postaci mieszanki zostanie poddane testowaniu, przy czym alternatywnymi specyfikacjami będą: standardowy model probitowy, model t Studenta o nieznannej liczbie stopni swobody oraz model oparty na modyfikacji rozkładu t Studenta o nieznannej liczbie stopni swobody, który dopuszcza asymetrię; zob. Marzec [2006], Osiewalski i Marzec [2004].

Zastosowanie mieszanki rozkładów zilustrujemy przykładem z bankowości. Model dwumianowy zostanie oszacowany na podstawie informacji o decyzjach kredytobiorców odnośnie spłaty kredytów detalicznych. Intencją autora jest zaprezentowanie możliwości wykorzystania tej klasy modeli dychotomicznych w scoringu kredytowym. Wydaje się to być interesująca, z punktu widzenia metodycznego, alternatywa dla takich tradycyjnych podejść jak model probitowy czy logitowy.

2. Definicja modelu i własności

Rozważany w artykule model dwumianowy ma postać

$$\begin{cases} z_t = \mathbf{x}_t \cdot \boldsymbol{\beta} + \varepsilon_t \\ y_t = I_{(0,\infty)}(z_t), \end{cases} \quad (1)$$

gdzie $I_{\Omega}(\omega)$ jest funkcją charakterystyczną zbioru, tzn. $I_{\Omega}(\omega) = 1$, gdy $\omega \in \Omega$ i $I_{\Omega}(\omega) = 0$, jeżeli $\omega \notin \Omega$, \mathbf{x}_t jest wektorem k zmiennych egzogenicznych (lub ich znanych funkcji) charakteryzujących obserwację o numerze t . Kluczowe założenie polega na przyjęciu niezależności identycznych rozkładów dla składników losowych ε_t , określonych przez skończoną mieszankę rozkładów normalnych. Rozkład nieobserwowalnej zmiennej z_t , która reprezentuje użyteczność podjętej decyzji, jest zdeterminowany przez rozkład zmiennej ε_t . W standardowym modelu probitowym czy logitowym zakłada się, że jest to rozkład symetryczny i jednomodalny. Natomiast w tym przypadku jego kształt może być całkowicie odmienny, np. wielomodalny lub asymetryczny.

Dla skończonych mieszanek funkcja gęstości i dystrybuanta jednowymiarowej zmiennej losowej ε_t jest określona przez wypukłą kombinację funkcji gęstości i dystrybuant składowych. W omawianym przypadku funkcja gęstości zmiennej losowej ε_t ma postać

$$p(\varepsilon_t | \boldsymbol{\theta}) = \sum_{j=1}^J \pi_j \cdot p_j(\varepsilon_t | \mu_j, \tau_j), \quad (2)$$

gdzie $J > 1$ jest liczbą składników mieszanki, $p_j(\cdot | \mu_j, \tau_j)$ jest funkcją gęstości zmiennej losowej o rozkładzie normalnym, o wartości oczekiwanej μ_j i precyzji (odwrotności wariancji) τ_j , π_j są wagami, tzn. $\pi_j \in (0;1)$ i $\sum_{j=1}^J \pi_j = 1$, $\boldsymbol{\theta}$ to wektor kolumna zawierający wszystkie powyższe parametry. Analogicznie, wartość dystrybuanty zmiennej ε_t w punkcie a wyraża formuła

$$F(a | \boldsymbol{\theta}) = \Pr(\varepsilon_t < a) = \sum_{j=1}^J \pi_j \cdot F_j(a | \mu_j, \tau_j), \quad (3)$$

gdzie $F_j(\cdot | \mu_j, \tau_j)$ oznacza wartość dystrybuanty zmiennej o rozkładzie normalnym, określonym przez parametry μ_j i τ_j . Badając własności tych rozkładów można wykorzystać fakt, że jak zauważył Pearson [1894], momenty zwykłe rzędu r zmiennej ε_t są liniową kombinacją momentów liczonych według rozkładów określonych przez funkcje gęstości poszczególnych składników mieszanki, tj.

$$E[\varepsilon_t^r] = \sum_{j=1}^J \pi_j \cdot E^{p_j}[\varepsilon_t^r]. \quad (4)$$

W konsekwencji zmienną losową określoną przez mieszankę rozkładów normalnych charakteryzuje to, że wszystkie jej momenty przyjmują skończone wartości, przy czym nieparzyste momenty centralne są równe zero. W szczególności wartość oczekiwana zmiennej ε_t jest równa $\mu_{mix} = \sum_{j=1}^J \pi_j \mu_j$, zaś wariancja $\tau_{mix}^{-1} = \sum_{j=1}^J \pi_j (\mu_j^2 + \tau_j^{-1}) - \mu_{mix}^2$. Na podstawie formuły (4) można wyprowadzić kurtozę i miarę asymetrii opartą na momencie centralnym trzeciego rzędu.

Atrakcyjność tej klasy rozkładów polega na tym, że można je stosować w zagadnieniach, w których jest potrzeba uwzględnienia rozkładów wielomodalnych, asymetrycznych lub wyostrzonych (eksces > 0) albo spłaszczonych (eksces < 0) w stosunku do rozkładu normalnego. Natomiast nie są one użyteczne przy modelowaniu zjawisk, których rozkłady charakteryzują się tzw. grubymi ogonami. W tym ostatnim przypadku zastosowanie ma rozkład t Studenta, który stanowi nieskończoną mieszankę rozkładów

normalnych względem parametru skali, gdy wagi mają rozkład gamma. Szerzej o wybranych modelach regresji uzyskiwanych za pomocą nieskończonych mieszanek rozkładów normalnych piszą Fernández i Steel [2000].

Istotną cechą modeli opartych na mieszkankach jest złożoność funkcji wiarygodności. Funkcja gęstości dla zmiennej ciągłej z_t ma postać

$$p(z_t|\boldsymbol{\theta}) = \sum_{j=1}^J \pi_j \cdot p(z_t|\mu_j, \tau_j) \quad \text{dla } t=1, \dots, T, \quad (5)$$

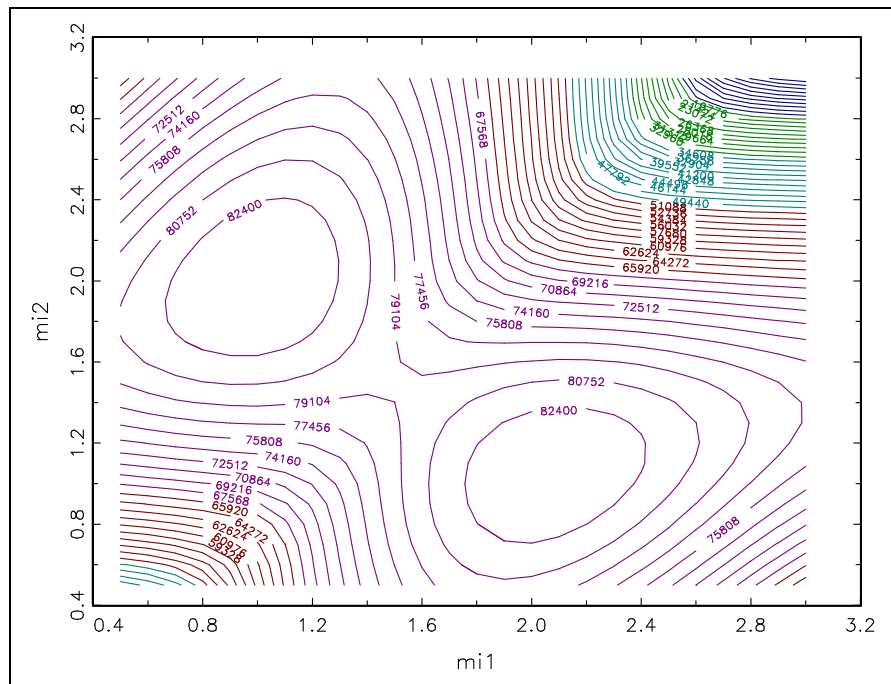
gdzie $\boldsymbol{\theta} = [\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_J]$ reprezentuje wektor nieznanych parametrów, $\boldsymbol{\theta}_j \in \Theta_j$ ($j=1, \dots, J$). W przypadku mieszanki normalnej mamy $\boldsymbol{\theta}'_j = [\mu_j, \tau_j, \pi_j]$. Problemy z estymacją wynikają z tego, że gdy $\Theta_1 = \dots = \Theta_J$, to rozkład dla z_t , $p(z_t|\boldsymbol{\theta})$, jest niezmienniczy ze względu na permutacje składników mieszanki, zob. np. Richardson i Green [1997], Celeux, Hurn i Robert [2000], Geweke [2007], Marin, Mengersen i Robert [2005]. W literaturze ten problem nosi nazwę *label switching problem* i wynika z prawa przemienności dodawania (wartość sumy nie zależy od porządku składników). Zatem kolejność składowych mieszanki może być dowolna, a przypisane numery mają charakter umowny. Negatywną tego konsekwencją jest wielomodalność funkcji wiarygodności, gdyż liczba symetrycznych ekstremów lokalnych jest równa liczbie permutacji zbioru J elementowego ($J!$). Badania symulacyjne pokazują, że im więcej obserwacji, tym wielomodalność jest bardziej wyraźna; zob. Geweke [2007]. Jest to zgodne z intuicją, gdyż wraz ze wzrostem liczby obserwacji funkcja wiarygodności staje się bardziej wyostrowiona w otoczeniu ekstremum (im mniej obserwacji, tym funkcja jest bardziej płaska). Zatem bez jakichkolwiek restrykcji na parametry, model oparty na mieszance rozkładów jest nieidentyfikowalny. Nawet w najprostszym przypadku próby prostej, gdy $J=2$ i parametry τ_j, π_j są ustalone, zaś estymacji podlegają parametry μ_1 i μ_2 , to uzyskuje się różne oceny MNW w zależności od doboru punktu startowego. Wielomodalność funkcji wiarygodności, uzyskanej na podstawie próby wygenerowanej przy założeniu $\mu_1=1$ i $\mu_2=2$, prezentuje wykres linii konturowych funkcji wiarygodności; zob. Rysunek 1. Wartości pozostałych parametrów ustaliliśmy następująco: $\pi_1 = 0,5$, $\tau_1 = 10$, $\tau_2 = 6.667$. Ekstremum znajduje się w dwóch punktach: $\mu_1=1$ i $\mu_2=2$ oraz $\mu_1=2$ i $\mu_2=1$, w których gęstość przyjmuje identyczną wartość. Wystarczającym warunkiem zapewniającym identyfikowalność parametrów jest narzucenie restrykcji na μ_j bądź τ_j . Zakłada się, że $\mu_{j-1} < \mu_j$, albo $\tau_{j-1} < \tau_j$, albo rzadziej $\pi_{j-1} < \pi_j$ dla $j=2, \dots, J$; zob. Geweke i Keane [1999], Koop [2005], Geweke [2007].

W przypadku analizowanego w artykule modelu dychotomicznego, gdy funkcja wiarygodności ma postać

$$p(\boldsymbol{\theta}; \mathbf{y}) = \prod_{t=1}^T (1 - F(-\mathbf{x}_t \boldsymbol{\beta} | \boldsymbol{\theta}))^{y_t} F(-\mathbf{x}_t \boldsymbol{\beta} | \boldsymbol{\theta})^{1-y_t}, \quad (6)$$

gdzie $F(-x_t \boldsymbol{\beta} | \boldsymbol{\theta})$ jest dane wzorem (3), wymaga się dodatkowo, aby jeden ze składników mieszanki miał ustaloną wartość parametru rozproszenia, tzn. aby istniało takie j^* , że np. $\tau_{j^*} = 1$; zob. Geweke i Keane [1999]. Jest to warunek niezbędny dla identyfikowalności funkcji wiarygodności, bo w przeciwnym przypadku istniałoby nieskończenie wiele kombinacji parametrów, dla których funkcja wiarygodności osiągałaby identyczne i maksymalne wartości. Wybór j^* jest dowolny.

Rysunek 1. Linie konturowe funkcji wiarygodności dla próby prostej, gdy $\mu_1=1$ i $\mu_2=2$.



Źródło: obliczenia własne.

Często rozważa się szczególne przypadki mieszanki rozkładów normalnych. W modelu z mieszanką względem średniej (ang. *mean mixture of normals*) zakłada się, że $\tau_j = 1$ dla każdego j oraz $\mu_{j-1} < \mu_j$ dla $j=2, \dots, J$. Zarówno w tym przypadku, jak i w modelu najogólniejszym wygodnie jest przyjąć, że brak sztucznej zmiennej „1”, gdyż w przeciwnym przypadku należałoby ustalić $\mu_{j^*}=0$ dla pewnego j^* . Natomiast w modelu z mieszanką względem parametru skali (ang. *scale mixture of normals*), gdy $\mu_j = 0$ dla każdego j i występuje wyraz wolny, to wymaga się, aby $\tau_{j-1} < \tau_j$ dla $j=2, \dots, J$. Gdyby w tym ostatnim przypadku nie byłoby wyrazu wolnego, to należałoby założyć, że jeden spośród parametrów μ_{j^*} jest swobodny i podlega estymacji. W niniejszym artykule podstawową specyfikacją jest model mieszanki zarówno względem parametru położenia, jak i parametru skali, więc przyjęliśmy, że w równaniu regresji dla zmiennej ukrytej z_i brak jest wyrazu wolnego.

W zdecydowanej części prac dotyczących mieszanek liczba ich składowych jest ustalona, najczęściej rozważa się przypadki $J = 2, 3$ lub 4 . Wówczas testowanie konkurencyjnych modeli można przeprowadzić poprzez tzw. bayesowskie kryterium informacyjne Schwarza; zob. Roeder i Wassermann [1997]. Na uwagę zasługuje praca Richardson i Green [1997], w której przyjęto, że liczba składowych jest wielkością nieznaną (losową), a następnie przedstawiono pełne wnioskowanie bayesowskie, łącznie z wyborem najlepszego modelu. Natomiast w tym opracowaniu rozważa się pełną mieszankę dwuskładnikową, zaś testowanie tej specyfikacji względem prostszych, tj. modelu mieszanki względem średniej czy precyzji, albo modelu podstawowego, będzie opierać się na prawdopodobieństwach a posteriori; zob. Geweke i Keane [1999], Qu i Qu [2000], Austin i Escobar [2003]. Szerzej o bayesowskim porównywaniu konkurencyjnych modeli piszą m.in. O’Hagan [1994], Osiewalski i Steel [1993], Kass i Raftery [1995], Osiewalski [2001].

3. Konstrukcja rozkładów a priori

Bayesowski model statystyczny jest jednoznacznie zdefiniowany poprzez łączny rozkład prawdopodobieństwa dla obserwacji i parametrów $p(\mathbf{y}, \boldsymbol{\theta}) = p(\mathbf{y}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})$, gdzie $p(\mathbf{y}|\boldsymbol{\theta})$ to rozkład próbkowy

warunkowy względem θ , zaś $p(\theta)$ to brzegowy rozkład parametrów zwany rozkładem a priori. Parametry są traktowane jako wielkości losowe, zaś wstępną o nich wiedzę reprezentuje właśnie rozkład a priori. Zasady bayesowskiej estymacji parametrów sprowadzają się do wyznaczenia z rozkładu łącznego, $p(y, \theta)$, warunkowej gęstości dla wektora parametrów θ przy danym y , danej wzorem Bayesa

$$p(\theta|y) = \frac{p(y, \theta)}{p(y)} \propto L(\theta; y) \cdot p(\theta), \quad (7)$$

gdzie $p(\theta; y)$ jest funkcją wiarygodności; zob. np. O'Hagan [1994], Osiewalski [2001], Koop [2003]. Rozkład a posteriori jest kombinacją rozkładów a priori i próbkowego, zatem niesie on - po zaobserwowaniu danych - pełną informację o interesujących charakterystykach opisujących zjawisko, w szczególności o parametrach. Wyznaczenie rozkładów a posteriori wg formuły (7) dla poszczególnych składowych wektora θ wymaga całkowania wielokrotnego, podobnie jak obliczenie wybranych charakterystyk tych rozkładów, np. wartości oczekiwanych i wariancji (jeżeli istnieją). W przypadku omawianych modeli realizuje się to wyłącznie za pomocą metod numerycznych.

W pierwszej kolejności omówimy zagadnienie konstrukcji rozkładów a priori. W przypadku modeli opartych na mieszankach wykluczone jest stosowanie niewłaściwych rozkładów a priori dla wszystkich parametrów składników mieszanki, μ_j i τ_j ($j=1, \dots, J$), gdyż prowadzi to do braku rozkładów a posteriori; zob. Diebolt i Robert [1994], Roeder i Wassermann [1997], Marin, Mengersen i Robert [2005]. W literaturze trwa dyskusja dotycząca struktury rozkładów a priori w kontekście konieczności uwzględnienia restrykcji dających identyfikowalność. Zgodność badaczy dotyczy przede wszystkim rozkładów dla parametrów β i π . W przypadku wektora β przyjmuje się zwykle k -wymiarowy rozkład normalny

$$p(\beta) = f_N^k(\beta | \underline{\beta}, \underline{H}_\beta^{-1}), \quad (8)$$

o wartości oczekiwanej a priori $\underline{\beta}$ i dodatnio określonej macierzy precyzji \underline{H}_β , które są ustalane przez badacza, zob. Geweke i Keane [1999].

Dla nieznanych wag stosuje się wielomianowy rozkład beta, czyli rozkład Dirichleta z parametrami $\underline{\pi}_1, \dots, \underline{\pi}_J$ ($\underline{\pi}_j > 0$ dla $j=1, \dots, J$), którego funkcja gęstości ma postać

$$p(\pi) = \Gamma(\underline{\pi}_1 + \dots + \underline{\pi}_J) [\Gamma(\underline{\pi}_1) \cdot \dots \cdot \Gamma(\underline{\pi}_J)]^{-1} (\pi_1)^{\underline{\pi}_1 - 1} \cdot \dots \cdot (\pi_J)^{\underline{\pi}_J - 1}, \quad (9)$$

gdzie $\Gamma(\cdot)$ oznacza funkcję gamma; zob. DeGroot [1981], Magiera [2005]. Brzegowe rozkłady dla π_j są rozkładami beta. Zwykle przyjmuje się $\underline{\pi}_j = 1$ dla $j=1, \dots, J$, więc każda składowa π_j ma brzegowy rozkład jednostajny na przedziale (0; 1).

W odniesieniu do pozostałych, kluczowych parametrów μ_j i τ_j , można w literaturze spotkać różne propozycje. Dyskusja skupia się na takim skonstruowaniu rozkładów a priori lub wprowadzeniu odpowiedniej reparametryzacji, aby możliwe było uwzględnienie restrykcji gwarantujących identyfikowalność parametrów i jednocześnie te rozkłady nie wносиły arbitralnej wiedzy, która nie znajduje potwierdzenia w danych. Zatem konstrukcja rozkładów a priori winna iść w parze z narzuceniem odpowiednich restrykcji gwarantujących identyfikowalność, $\tau_{j-1} < \tau_j$, albo $\mu_{j-1} < \mu_j$, albo $\pi_{j-1} < \pi_j$ ($j=2, \dots, J$). Narzucenie restrykcji nie jest łatwe. Po pierwsze, ich uwzględnienie powoduje problemy numeryczne, co podnoszą Celeux, Hurn i Robert [2000]. Po drugie, szczególnie trudno jest dobrać rozkłady dla parametrów μ_j , gdyż te przyjmują wartości z całego zbioru liczb rzeczywistych. A jest to konieczne, gdy rozważa się model z mieszanką rozkładów względem średniej. Pewne rozwiązanie w tym zakresie proponuje Mengersen i Robert [1993] oraz Roeder i Wassermann [1997], proponując tzw. rozkłady a priori typu Markowa (ang. *Markov prior*).

Dla parametrów precyzji τ_j przyjmuje się a priori rozkłady z rodziny gamma, np. chi kwadrat lub wykładniczy; zob. Geweke i Keane [1999] oraz Roeder, Wassermann [1997], Austin i Escobar [2003], Koop [2003]. Celeux, Hurn i Robert [2000] stosuje rozkład wykładniczy dla odchylenia standardowego ($\tau_j^{-0.5}$), co prowadzi do rozkładu dla precyzji τ_j charakteryzującego się nieskończoną wartością oczekiwaną i brakiem

momentów wyższych rzędów. W przypadku wektora $\boldsymbol{\mu}' = [\mu_1 \dots \mu_J]$ najprościej jest przyjąć J -wymiarowy rozkład normalny. Zatem, gdy restrikcje narzuca się na parametry precyzji, to $p(\boldsymbol{\mu}) = f_N^J(\boldsymbol{\mu} | \underline{\boldsymbol{\mu}}, \underline{\mathbf{H}}_{\boldsymbol{\mu}}^{-1})$, jak proponuje Geweke i Keane [1999]. Natomiast u Koopa [2005] restrikcje dotyczą $\boldsymbol{\mu}$, więc $p(\boldsymbol{\mu}) \propto f_N^J(\boldsymbol{\mu} | \underline{\boldsymbol{\mu}}, \underline{\mathbf{H}}_{\boldsymbol{\mu}}^{-1}) \cdot I(\mu_1 < \mu_2 < \dots < \mu_J)$, co prowadzi do wielowymiarowego, uciętego rozkładu normalnego, przy czym $\underline{\mathbf{H}}_{\boldsymbol{\mu}}$ nie musi być macierzą diagonalną.

W tym artykule rozważamy mieszanekę dwuskładnikową ($J=2$), więc narzucenie restrikcji nie nastęrcza problemów, gdy uwzględni się je w rozkładzie a priori dla precyzji; podobnie uczynili Geweke i Keane [1999]. Można je narzucić na dwa sposoby, uzyskując równoważne parametryzacje. Niech wg pierwszej parametryzacji $\tau_1 = 1$ i $\tau_2 > 1$, gdy pozostałymi parametrami mieszanki są μ_1 , μ_2 i π_1 . Zatem w drugim przypadku $\tau_1^* < 1$ i $\tau_2^* = 1$, przy czym parametry obu specyfikacji są związane zależnością: $\tau_1^* = 1/\tau_2$, $\mu_1^* = \mu_2/\tau_2$, $\mu_2^* = \mu_1/\tau_2$ i $\pi_1^* = 1 - \pi_1$. W tym artykule zakładamy, że $\tau_1 = 1$ i $\tau_2 > 1$, więc dla ostatniego parametru jest wygodnie przyjąć lewostronnie ucięty rozkład wykładniczy z parametrem $r_2 > 0$.¹ W konsekwencji warunkowa wartość oczekiwana w rozkładzie a priori, $E[\tau_2 | \tau_2 > 1]$, wynosi $1 + r_2$, zaś wariancja $(r_2)^2$. W przypadku drugiej parametryzacji, $\tau_1^* < 1$ i $\tau_2^* = 1$, stosując przekształcenie zmiennych losowych konstruujemy rozkłady dla τ_2^* , μ_1^* , μ_2^* i π_1^* . Takie spójne podejście gwarantuje, że rozkłady a posteriori dla obu specyfikacji także będą równoważne. Przyjęcie dla τ_2 uciętego rozkładu wykładniczego powoduje, iż dla τ_1^* uzyskamy odwrotny rozkład gamma określony na przedziale (0; 1). Jego funkcja gęstości $p(\tau_1^*)$ jest ograniczona od góry, ponadto dla $r_2 > 0.5$ posiada modalną w punkcie $(2 \cdot r_2)^{-1}$, zaś dla $0 < r_2 < 0.5$ jest wyłącznie funkcją rosnącą. Wartość stałej r_2 można ustalić na poziomie 1, 2, 3 lub 4.

W prezentowanych dalej badaniach w odniesieniu do stałych definiujących rozkłady a priori przyjęto: $r_2 = 1$ w przypadku rozkładu dla precyzji τ_2 , $\underline{\boldsymbol{\beta}} = \mathbf{0}_{[k \times 1]}$ i $\underline{\mathbf{H}}_{\boldsymbol{\beta}} = \mathbf{I}_{[k \times k]}$ dla $\boldsymbol{\beta}$, $\underline{\boldsymbol{\mu}} = \mathbf{0}_{[2 \times 1]}$ i $\underline{\mathbf{H}}_{\boldsymbol{\mu}} = \mathbf{I}_{[2 \times 2]}$ dla $\boldsymbol{\mu}$. Wstępną informację o wadze $\pi_1 \in (0; 1)$ reprezentował rozkład jednostajny. Przeprowadzając analizę wrażliwości można ex post określić wpływ stałych definiujących rozkłady a priori na kształt rozkładów a posteriori.

4. Uzyskiwanie rozkładów a posteriori

W celu uzyskania rozkładów a posteriori i ich charakterystyk wykorzystaliśmy algorytm Metropolisa i Hastingsa. Wygodnie jest ze względów numerycznych wprowadzić taką parametryzację, aby zbiory dopuszczalnych wartości dla wszystkich parametrów były zbiorami liczb rzeczywistych. Przyspiesza to numeryczną realizację algorytmu. Gdy $J=2$, to wymiar wektora $\boldsymbol{\theta}' = [\boldsymbol{\beta}' \boldsymbol{\mu}' \theta_r \theta_\pi]$ wynosi $k+4$. Dla wagi π_1 zastosowaliśmy transformację logistyczną, więc $\theta_\pi = \ln(\pi_1 / (1 - \pi_1))$ ma standardowy rozkład logistyczny; zob. także Celeux, Hurn i Robert [2000]. Gdy narzuca się restrikcję $\tau_1 = 1$ i $\tau_2 > 1$, to wygodnie jest zdefiniować zmienną $\theta_r = \ln(\tau_2 / r_2)$, zaś gdy $\tau_1^* < 1$ i $\tau_2^* = 1$ przyjąć $\theta_r = -\ln(\tau_1^* / r_2)$. Wówczas θ_r ma ucięty na lewo od wartości $-\ln(r_2)$ standardowy rozkład Gumbela.

Uwzględniając płynące z literatury sugestie na temat konstrukcji rozkładów a priori, łączny rozkład a priori dla wektora wszystkich parametrów modelu ma postać

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\beta}) \cdot p(\boldsymbol{\mu}) \cdot p(\theta_r) \cdot p(\theta_\pi), \quad (10)$$

¹ Wówczas uzyskamy ujemny rozkład wykładniczy z parametrem położenia równym 1 i parametrem skali r_2 ; zob. Magiera [2005].

gdzie $p(\theta_\tau) = \exp(-\exp(\theta_\tau)) \cdot \exp(\theta_\tau + r_2^{-1}) \cdot I_{(-\ln r_2, +\infty)}(\theta_\tau)$, $p(\boldsymbol{\beta}) = f_N^k(\boldsymbol{\beta} | \underline{\boldsymbol{\beta}}, \underline{\mathbf{H}}_\beta^{-1})$,
 $p(\boldsymbol{\mu}) = f_N^2(\boldsymbol{\mu} | \underline{\boldsymbol{\mu}}, \underline{\mathbf{H}}_\mu^{-1})$, $p(\theta_\pi) = \exp(\theta_\pi) / (1 + \exp(\theta_\pi))^2$.

Łączny rozkład dla obserwacji \mathbf{y} i parametrów $\boldsymbol{\theta}$, stanowiący nasz model bayesowski, ma postać

$$p(\mathbf{y}, \boldsymbol{\theta}) = p(\mathbf{y} | \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}) \cdot \prod_{t=1}^T (1 - F(-\mathbf{x}_t \boldsymbol{\beta} | \boldsymbol{\theta}))^{y_t} F(-\mathbf{x}_t \boldsymbol{\beta} | \boldsymbol{\theta})^{1-y_t}. \quad (11)$$

W celu praktycznej realizacji wnioskowania bayesowskiego zastosowano algorytm Metropolis i Hastingsa. Za pomocą tego mechanizmu losowania uzyskaliśmy wektory liczb pseudolosowych $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(n)}$ z rozkładu a posteriori $p(\boldsymbol{\theta} | \mathbf{y})$. Rozkłady a posteriori aproksymuje się histogramami skonstruowanymi na podstawie tej próby pseudolosowej. W szczególności, w celu wyznaczenia wartości oczekiwanej wg rozkładu a posteriori pewnej skalarnej funkcji wektora $\boldsymbol{\theta}$, $g(\boldsymbol{\theta})$, czyli aproksymacji całki $E[g(\boldsymbol{\theta}) | \mathbf{y}] = \int g(\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}$, oblicza się średnią empiryczną

$$\frac{1}{n} \sum_{i=1}^n g(\boldsymbol{\theta}^{(i)}). \quad (12)$$

Szczegółowe omówienie zastosowanej procedury Monte Carlo opartej na łańcuchach Markowa (MCMC) można znaleźć w pracach: O'Hagan [1994], Tierney [1994], Gamerman [1997] lub Koop [2003], także Magiera [2005]. W prezentowanych badaniach za funkcję pomocniczą, z której losuje się liczby pseudolosowe, przyjęto wielowymiarowy rozkład t Studenta o 3 stopniach swobody, modalnej równej poprzedniemu wylosowanemu wektorowi $\boldsymbol{\theta}^{(i-1)}$ oraz ustalonej wcześniej macierzy precyzji; zob. Osiewalski i Marzec [2004]. Jako wartości startowe dla parametrów przyjęto: dla $\boldsymbol{\beta}$ - bayesowskie oceny modelu probitowego, zaś dla pozostałych: $\pi_1 = 0,5$, $\mu_1 = -0,1$, $\mu_2 = 0,1$, $\tau_2 = 1,1$.

5. Wyniki empiryczne

W tym rozdziale prezentujemy wyniki bayesowskich modeli dwumianowych z dwuelementową mieszanką rozkładów normalnych. W tym celu wykorzystaliśmy informację o niespłacalności kredytów detalicznych. Zbiór danych obejmował 39034 rachunków kredytowych, pochodził z lat 2000-2001 i był wykorzystywany w innych pracach autora, zob. Marzec [2006]. Przyjęto, że $y_t = 1$, jeżeli kredyt jest zakwalifikowany do kategorii zagrożonych (poniżej standardu, wątpliwych i straconych). W przeciwnym przypadku – dla kredytów z grupy należności normalnych – $y_t = 0$. Zbiór potencjalnych zmiennych egzogenicznych wyjaśniających ryzyko pojedynczej umowy kredytowej zawierał: płeć, wiek kredytobiorcy, wpływy na rachunki ROR, posiadanie rachunku ROR, informację o tym, czy kredytobiorca posiada karty płatnicze lub kredytowe wydane przez ten bank, sposób udzielenia kredytu (przez pośrednika kredytowego albo bezpośrednio przez bank), typ kredytu (kredyt konsumpcyjny albo hipoteczny), okres trwania umowy kredytowej, kwota przyznanego kredytu, waluta kredytu, podstawowe źródło dochodu uzyskiwanego przez kredytobiorcę (umowa o pracę, albo renta lub emerytura, albo własna działalność, umowa o dzieło lub umowa zlecenie, albo inne źródło). W konsekwencji liczba składowych wektora $\boldsymbol{\beta}$ wynosiła $k=13$ (bez wyrazu wolnego).

5.1. Porównywanie konkurencyjnych modeli

Proponując model dwumianowy z mieszanką rozkładów normalnych (model M_1) można postawić pytanie, czy w świetle posiadanych danych takie uogólnienie standardowego modelu probitowego (M_4) było zasadne? Jeżeli tak, to czy model M_1 lepiej „pasuje” do danych, niż inne uogólnienia probitu?

Rozważając dwumianowy model postaci (1) i (2) zakładamy, że rozkład składnika losowego może być asymetryczny, wielomodalny lub wyostrzony albo spłaszczony. Innymi, prostszymi specyfikacjami są: mieszanka względem średniej ($\tau_j = 1$ dla $j=1, \dots, J$; model M_2) albo precyzji ($\mu_j = 0$ dla $j=1, \dots, J$; model M_3). W pierwszym przypadku rozkład składnika ε_i może posiadać rozkład o własnościach jak w M_1 , zaś w drugim mamy do czynienia wyłącznie z rozkładem symetrycznym i leptokurtycznym. Modele M_2 , M_3 i M_4 uzyskujemy poprzez warunkowanie względem parametrów w M_1 . Zatem ich rozkłady a priori otrzymaliśmy poprzez narzucenie odpowiednich restrykcji na parametry w rozkładzie a priori modelu M_1 . Dodatkowo w porównaniu uwzględniliśmy wyniki innych modeli uogólniających model probitowy. Rozważyliśmy model t Studenta o nieznannej liczbie stopni swobody (M_5), w którym dodatkowo zależność między z_i a zmiennymi egzogenicznymi jest w formie wielomianu drugiego stopnia; zob. Marzec [2006]. Rozkład t Studenta można traktować jako nieskończoną mieszankę względem precyzji (M_3), która dopuszcza grube ogony rozkładu dla ε_i , ale narzuca jednomodalność i symetrię. Zatem model M_5 jest uogólnieniem specyfikacji M_3 . Inną interesującą alternatywą wydaje się być model z modyfikacją rozkładu t Studenta, który dopuszcza asymetrię (M_6), zob. Osiewalski i Marzec [2004]. Należy podkreślić, że modele M_1 i M_6 są specyfikacjami nie zagnieżdżonymi, więc ich testowanie byłoby trudne na gruncie klasycznym, w przeciwieństwie do wnioskowania bayesowskiego, które daje nam do tego proste narzędzie.

Porównanie modeli przeprowadzono w oparciu o prawdopodobieństwo a posteriori liczone wg formuły (zob. np. Osiewalski [2001], Koop [2005])

$$p(M_i | \mathbf{y}) = \frac{p(\mathbf{y} | M_i) \cdot p(M_i)}{\sum_{h=1}^6 p(\mathbf{y} | M_h) \cdot p(M_h)} \quad \text{dla } i \in \{1, \dots, 6\}, \quad (13)$$

gdzie $p(M_i)$ (dla $i=1, \dots, 6$) oznacza prawdopodobieństwo a priori poszczególnych specyfikacji, zaś $p(\mathbf{y} | M_h)$ jest brzegową gęstością obserwacji w h -tym modelu, liczoną w oparciu o średnią harmoniczną wartości wiarygodności, jak to proponują Newton i Raftery [1994].

Z uwagi na fakt, że badane modele posiadają różną liczbę parametrów (k'_i), przy ustalaniu ich prawdopodobieństw a priori przyjęto zasadę, że preferowane są oszczędne parametryzacje; zob. Osiewalski i Steel [1993], Osiewalski [2001]. Najprostszemu modelowi probitowemu przypisano największe prawdopodobieństwa a priori równe 0,47. Model M_1 posiada prawdopodobieństwo równe 0,06 i jest ono dwukrotnie mniejsze niż prawdopodobieństwo modelu mieszanki względem średniej (M_2) i skośnego t Studenta, a nawet trzykrotnie mniejsze od mieszanki względem precyzji (M_3). Szczegółowe wyniki dotyczące mocy wyjaśniającej poszczególnych modeli prezentuje Tabela 1.

Tabela 1. Brzegowe gęstości wektora obserwacji i prawdopodobieństwa a posteriori badanych modeli

Model	Liczba parametrów	$\ln p(\mathbf{y} M_i)$	$p(M_i) \propto 2^{-k'_i}$	$p(M_i \mathbf{y})$
M_1	17	-13569	0,06	$1,4 \times 10^{-10}$
M_2 ; $\tau_1 = \tau_2 = 1$	16	-13836	0,12	$3,1 \times 10^{-126}$
M_3 ; $\mu_1 = \mu_2 = 0$	15	-13845	0,24	$7,6 \times 10^{-130}$
M_4 : probitowy	14	-13846	0,47	$5,6 \times 10^{-130}$
M_5 : t-Studenta	80	-13600	$6,38 \times 10^{-21}$	$5,2 \times 10^{-43}$
M_6 : skośny t-Studenta	16	-13547	0,12	≈ 1

Źródło: obliczenia własne.

Porównując model M_1 dla tle M_2 , M_3 i M_4 widzimy, że jest on zdecydowanie najbardziej prawdopodobny. Podobnie jest w odniesieniu do M_5 , który dopuszcza wyłącznie grube ogony i wyostrzenie rozkładu. Szczegółowa analiza wyników prowadzi nas do konkluzji, iż model M_3 , który zakłada leptokurtyczność w stosunku do rozkładu normalnego, w świetle tych danych niewiele wnosi w stosunku do probitu (M_4). Dane skłaniają się ku rozkładom asymetrycznym, gdyż prawdopodobieństwo modelu M_2 jest większe niż M_3 i M_4 . Pierwszy model jest cztery rzędy wielkości bardziej prawdopodobny niż dwa pozostałe. Jednocześnie dane preferują rozkład o grubych ogonach, bo model M_5 jest co prawda gorszy od

M_1 , ale bardziej prawdopodobny niż M_2 , M_3 i M_4 . Modelem, który łączy asymetrię z grubymi ogonami, jest M_6 . Jego prawdopodobieństwa wynosi prawie jeden i jest dziesięć rzędów wielkości bardziej prawdopodobny niż następny w rankingu model z pełną mieszanką (M_1). Wobec powyższego dane przedkładają skośny rozkład z grubymi ogonami nad rozkład ze skończonymi momentami wszystkich rzędów i ewentualną asymetrią lub dwumodalnością. Model mieszanki dwuelementowej jest interesujący, ale wyłącznie w kontekście pozostałych specyfikacji. Alternatywą dla M_6 może być trzelementowa mieszanka, lecz w stosunku do M_1 nie zmienia ona grubości ogonów i ma o trzy parametry więcej. Innym kierunkiem dalszych badań mogłaby być konstrukcja skończonej mieszanki rozkładów t Studenta o nieznanym liczbach stopni swobody, która pozwoliłaby uzyskać asymetryczny rozkład wielomodalny z grubymi ogonami.

5.2. Wyniki estymacji

Poniżej prezentujemy szczegółowe wyniki estymacji parametrów modeli opartych na mieszankach. Uwagę zwraca zgodność znaków ocen punktowych parametrów przy zmiennych objaśniających we wszystkich modelach (wyjątek to parametr przy zmiennej ROR). Zatem znaki efektów krańcowych w tych modelach są identyczne. Przeważająca większość parametrów jest statystycznie istotna, tzn. bayesowskie przedziały o największej gęstości a posteriori – odpowiedniki klasycznych przedziałów ufności – nie zawierają wartości zero. W modelu M_1 płęć, kwota i typ kredytu (konsumpcyjny, hipoteczny) wydają się nie mieć istotnego wpływu na prawdopodobieństwo niespłacenia kredytu. Odchylenia standardowe a posteriori dla wszystkich parametrów β były zdecydowanie mniejsze od jedności, czyli odchylenia standardowego a priori. Zatem rozkłady a posteriori charakteryzują się mniejszym rozproszeniem niż rozkłady a priori. Wyniki mieszanek względem średnich (M_2) i precyzji (M_3) są bardzo zbliżone do ocen modelu probitowego.

Przypomnijmy, iż parametry μ_1 , μ_2 i τ_2 nie mogą być bezpośrednio interpretowalne w modelu M_1 , gdyż jak wcześniej wspomniano, w mieszance dwuskładnikowej z restrykcjami dotyczącymi precyzji istnieją dwie równoważne parametryzacje: $\tau_1 = 1$ i $\tau_2 > 1$ oraz $\tau_1^* = 1/r_2 < 1$ i $\tau_2^* = 1$. Natomiast można porównywać ilorazy parametrów τ_i/τ_j i μ_i/τ_j . Wszystkie kluczowe parametry definiujące rozważane mieszanki (μ_1 , μ_1 , τ_2 , π_1) są statystycznie istotne. W modelu probitowym zakłada się, że wszystkie obserwacje pochodzą z jednej homogenicznej próby, natomiast w modelach dwuelementowej mieszanki przyjmuje się, że istnieją dwie odrębne podpróby. O podziale próby informuje waga π_1 , której wartość oczekiwana a posteriori w modelu M_1 wynosi 0,01, co wskazuje na nierówny podział zbioru obserwacji. Te same informacje uzyskujemy w modelu mieszanki względem średniej M_2 , w którym przyjęto $\tau_1 = \tau_2 = 1$.

W modelu probitowym wyraz wolny można traktować jako ucięcie, którego wartość decyduje o tym, czy zaobserwowano $y_i = 1$ czy $y_i = 0$. Ten parametr jest nieidentyfikowalny, gdy swobodnym parametrem jest wartość oczekiwana rozkładu dla ε_t , μ_{mix} . Wówczas μ_{mix} przejmują rolę wyrazu wolnego. W ramach specyfikacji M_2 , M_3 i M_4 można porównywać μ_{mix} i precyzję rozkładu dla ε_t , czyli τ_{mix} . W modelu M_2 ocena a posteriori dla μ_{mix} wynosi $-1,27 (\pm 0,14)$, natomiast w M_3 otrzymujemy, iż $E[\mu_{mix} | \mathbf{y}] = -1,125 (\pm 0,130)$, gdy tymczasem w modelu probitowym $E[\mu_{mix} | \mathbf{y}] = -1,220 (\pm 0,124)$. Wartości oczekiwane a posteriori dla precyzji τ_{mix} wynoszą odpowiednio: $E[\tau_{mix} | \mathbf{y}] = 1,04 (\pm 0,10)$ w M_2 oraz $0,85 (\pm 0,07)$ w M_3 . W modelu M_4 precyzja jest ustalona na poziomie 1, co wynika ze znanego warunku identyfikowalności. Wszystkie te rezultaty są zbliżone do siebie.

Tabela 2. Wartości oczekiwane i odchylenia standardowe parametrów poszczególnych modeli (zał. $\tau_1 = 1$ i $\tau_2 > 1$)

	Model M_1		Model M_2		Model M_3		Model M_4	
Zmienna (parametr)	$E(\cdot \mathbf{y})$	$D(\cdot \mathbf{y})$	$E(\cdot \mathbf{y})$	$D(\cdot \mathbf{y})$	$E(\cdot \mathbf{y})$	$D(\cdot \mathbf{y})$	$E(\cdot \mathbf{y})$	$D(\cdot \mathbf{y})$
„1”	-	-	-	-	-1,125	(0,130)	-1,220	(0,124)
Płęć	0,005	(0,004)	0,037	(0,018)	0,029	(0,017)	0,038	(0,018)

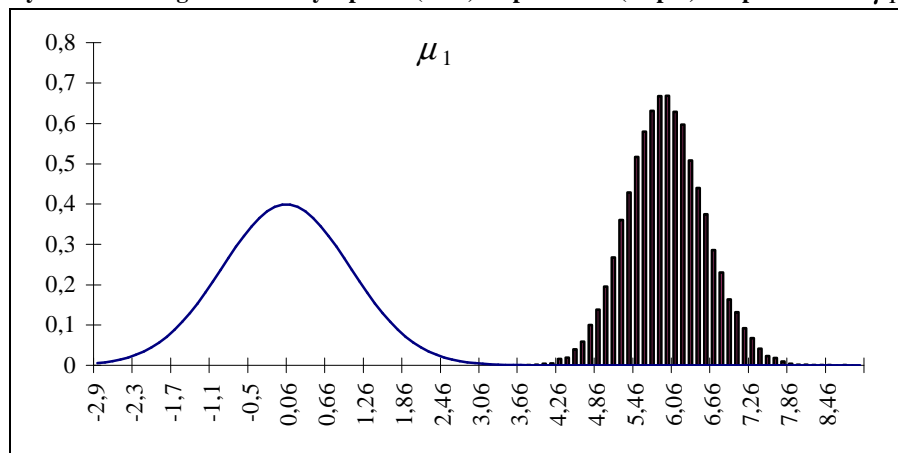
Wiek	-0,166	(0,022)	-0,886	(0,09)	-0,804	(0,075)	-0,866	(0,085)
Wpływy	-4,808	(0,336)	-2,370	(0,234)	-1,749	(0,176)	-1,731	(0,175)
ROR	0,081	(0,011)	-0,282	(0,038)	-0,260	(0,042)	-0,294	(0,038)
Karty	-0,024	(0,009)	-0,174	(0,035)	-0,157	(0,032)	-0,169	(0,033)
Pośrednik	0,243	(0,019)	1,325	(0,036)	1,199	(0,052)	1,292	(0,032)
Typ kredytu	0,173	(0,100)	0,289	(0,132)	0,233	(0,114)	0,247	(0,114)
Okres	-0,089	(0,015)	-0,272	(0,062)	-0,242	(0,058)	-0,237	(0,057)
Kwota	0,019	(0,010)	0,117	(0,032)	0,104	(0,028)	0,117	(0,029)
Waluta	0,152	(0,102)	0,402	(0,153)	0,347	(0,135)	0,331	(0,136)
Zrdoch1	-0,023	(0,006)	-0,097	(0,031)	-0,082	(0,026)	-0,094	(0,029)
Zrdoch2	0,030	(0,011)	0,307	(0,043)	0,287	(0,038)	0,304	(0,040)
Zrdoch3	-0,076	(0,017)	-0,200	(0,08)	-0,168	(0,080)	-0,189	(0,077)
μ_1	5,916	(0,604)	-1,290	(0,144)	-	-	-	-
μ_2	-0,343	(0,103)	0,872	(0,186)	-	-	-	-
τ_2	28,359	(4,179)	-	-	1,473	(0,216)	-	-
π_1	0,011	(0,001)	0,990	(0,004)	0,523	(0,148)	-	-

Źródło: obliczenia własne.

Zbadaliśmy także własności rozkładu składnika losowego w modelu M_1 , aby sprawdzić w jakim stopniu różni się od własności standaryzowanego rozkładu normalnego. W tym celu wykorzystaliśmy, jak sugeruje Geweke [2007], eksces i współczynnik asymetrii. Obie miary są funkcjami jednorodnymi stopnia zero ze względu na μ_j i τ_j (dla $j=1, \dots, J$). Obliczanie wartości oczekiwanej i wariancji nie jest zasadne, gdyż te charakterystyki nie są identyfikowalne. Wartość oczekiwana ekscesu wynosi 80,6, a odchylenie standardowe $\pm 6,4$, zaś współczynnik asymetrii jest dodatni i wynosi 8,6 ($\pm 0,32$). Wyniki wskazują, że mamy do czynienia z rozkładem prawostronnie asymetrycznym, który w stosunku do rozkładu normalnego jest bardzo wyostrojony wokół wartości modalnej.

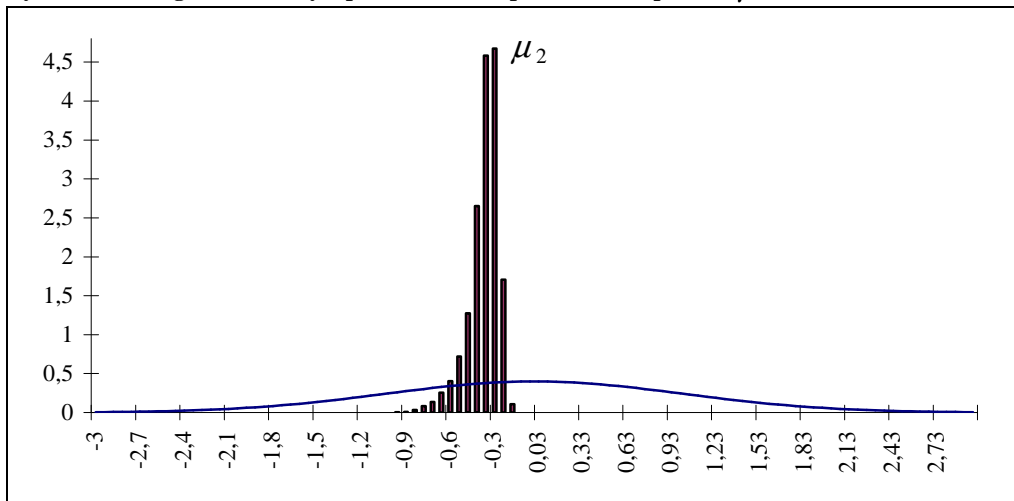
W uzupełnieniu prezentujemy histogramy rozkładów a priori i a posteriori dla μ_1 , μ_2 i τ_2 . Jako rozkład a priori przyjęliśmy dla μ_1 i μ_2 standaryzowany rozkład normalny, zaś dla precyzji (τ_2) ucięty rozkład wykładniczy o wartości oczekiwanej równej 2 i jednostkowym odchyleniu standardowym. Rozkład a posteriori dla μ_1 jest silnie przesunięty na prawo w stosunku do rozkładu a priori i charakteryzuje się większą precyzją, więc jest bardziej skupiony wokół modalnej. Natomiast w przypadku μ_2 obserwujemy, że rozkład a posteriori jest lewostronnie asymetryczny, jego wartość modalna a posteriori wynosi -0,27, a ponadto charakteryzuje się on bardzo małym rozproszeniem w stosunku do rozkładu a priori. W przypadku precyzji τ_2 widzimy, że histogram rozkładu a posteriori jest co prawda bardzo rozproszony, ale znajduje się w obszarze o bardzo małym prawdopodobieństwie a priori. Rola danych w stosunku do założeń a priori była decydująca w kształtowaniu wyników a posteriori.

Rysunek 2. Brzegowe rozkłady a priori (linia) i a posteriori (słupki) dla parametrów μ_1 i μ_2 .



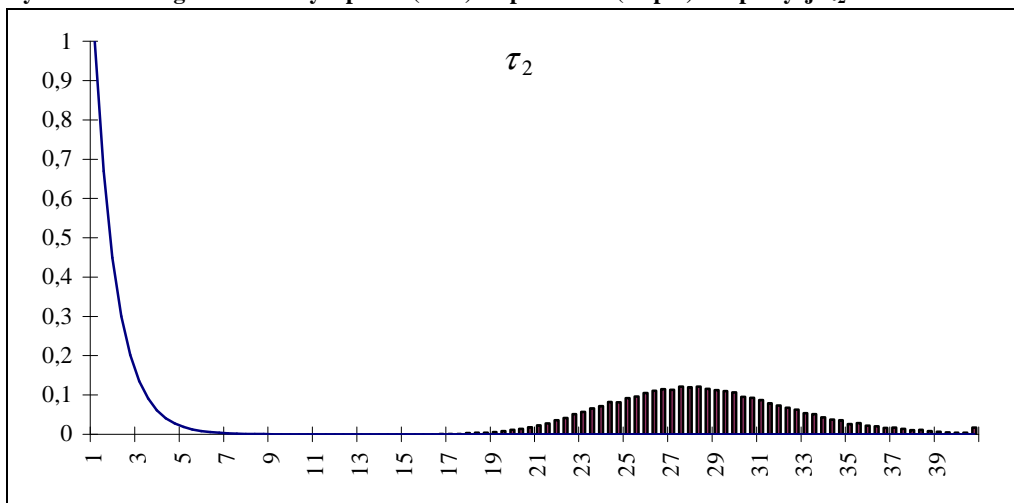
Źródło: obliczenia własne.

Rysunek 3. Brzegowe rozkłady a priori (linia) i a posteriori (słupki) dla μ_2 .



Źródło: obliczenia własne.

Rysunek 4. Brzegowe rozkłady a priori (linia) i a posteriori (słupki) dla precyzji τ_2 .



Źródło: obliczenia własne.

Podsumowanie

W niniejszym artykule rozważano model dwumianowy oparty na skończonej mieszance rozkładów normalnych jako uogólnienie modelu probitowego. Przedstawiono podstawowe własności mieszanek, zwracając uwagę na problem wielomodalności funkcji wiarygodności. Następnie omówiono specyfikację modelu bayesowskiego, w tym konstrukcję rozkładów a priori w kontekście identyfikowalności parametrów i skutecznego wykorzystania metod MCMC (algorytmu Metropolisa i Hastingsa).

W warstwie empirycznej dokonano bayesowskiej estymacji modeli opartych na mieszankach, wykorzystując dane o spłacalności kredytów. Następnie porównano konkurencyjne specyfikacje wykorzystując prawdopodobieństwa a posteriori. Celem artykułu było także zbadanie, czy w świetle posiadanych danych uzasadnione było zastosowanie mieszanek jako uogólnienia modelu probitowego. A jeżeli tak, to czy lepiej one opisują badane zjawisko niż modele dwumianowe z rozkładem t Studenta. Wyniki a posteriori pokazały, że dwuskładnikowa mieszanka posiada większe prawdopodobieństwo niż model probitowy, a także model zakładający rozkład t Studenta (symetryczny) z grubymi ogonami. Jednakże w porównaniu ze specyfikacją dopuszczającą asymetrię i grube ogony model mieszanki został negatywnie zweryfikowany. Posiadane dane preferują modele z rozkładem asymetrycznym i jednocześnie grubymi

ogonami, ale nie potwierdzają potrzeby wprowadzenia wielomodalności. Wnioski te wskazują na kierunek dalszych badań, który mógłby dotyczyć mieszanek opartych na rozkładach t Studenta.

Bibliografia

- Aitkin M. [2001], Likelihood and Bayesian Analysis of Mixtures, Statistical Modeling, nr 1, s. 287-304.
- Allenby G.M., R.P. Leone, L.C. Jen [1999], A Dynamic Model of Purchase Timing with Application to Direct Marketing, Journal of the American Statistical Association, vol. 94, nr 446; s. 365-374.
- Amemiya T. [1985], Advanced Econometrics, Harvard University Press, Cambridge (Massachusetts).
- Austin P., M. Escobar [2003], The Use of Finite Mixture Models to Estimate the Distribution of the Health Utilities Index in the Presence of a Ceiling Effects, Journal of Applied Statistics, vol. 30, nr 8, s. 909-923.
- Celeux G., M. Hurn, C.P. Robert [2000], Computational and Inferential Difficulties With Mixture Posteriori Distribution, Journal of the American Statistical Association, vol. 95, nr 451; s. 957-970.
- DeGroot M. [1981], Optymalne decyzje statystyczne, PWN, Warszawa.
- Dempster A.P., N.M. Laird, D.B. Rubin [1977], Maximum Likelihood From Incomplete Data via the EM Algorithm (with discussion), Journal of the Royal Statistical Society B, vol. 39, s. 1-38.
- Diebolt J., C.P. Robert [1994], Estimation of Finite Mixture Distributions though Bayesian Sampling, Journal of the Royal Statistical Society B, vol. 56, nr 2, s. 363-375.
- Dueker M. [1999], Conditional Heteroscedasticity in Qualitative Response Models of Time Series, A Gibbsa –Sampling Approach in the Bank Prime Rate, Journal of Business and Economic Statistics, vol 17, nr 4, s. 466-472.
- Erkanli A., D. Stangl, P. Müller [1993], A Bayesian Analysis of Ordinal Data Using Mixtures, Technical Report 93-01 Institute of Statistics and Decision Sciences, Duke University.
- Fernández C., M.F.J Steel [2000], Bayesian Regression Analysis with Scale Mixture of Normals, Econometric Theory, 16, s. 80-101.
- Frühwirth-Schnatter S. [2001], Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models, Journal of the American Statistical Association, vol. 96, nr 453; s. 194-209.
- Frühwirth-Schnatter S., R. Frühwirth [2006], Auxiliary Mixture Sampling with Applications to Logistics Models, Computational Statistics and Data Analysis, vol. 51, nr 7.
- Gamerman D. [1997], Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference, Chapman and Hall, London.
- Geweke J. [2007], Interpretation and Inference in Mixture Models: Simple MCMC Works, Computational Statistics and Data Analysis, vol. 51, nr 7, s. 3529-3550.
- Geweke J., M. Keane [1999], Mixture of Normals Probit Models, [w:] Analysis of Panel and Limited Dependent Variables: A Volume in Honor of G.S. Maddala, red. Hsiao C., Lahiri K., Lee L.F., Pesaran M.H., Cambridge University Press, Cambridge.
- Hamilton J.D. [1989], A New Approach to the Economic Analysis of Nonstationary time Series and the Business Cycle, Econometrica, 57, s. 357-384.
- Jasra A., C.C. Holmes, D.A. Stephens [2005], Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modeling, Statistical Science, vol. 20, nr 1, s. 50-67.
- Jedidi K., H.S. Jagpal. W.S. Desarbo [1997], Finite Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity, Marketing Science, 16, s. 35-859.
- Kass R.E., A. Raftery [1995], Bayes Factor, Journal of the American Statistical Association, vol. 90, nr 90, s. 773-795.
- Koop G. [2003], Bayesian Econometrics, Wiley, Chichester.
- Lesage J.P. [1992], Scoring the Compositing Leading Indicators – a Bayesian Turning Points Approach, Journal of Forecasting, 11, s. 35-46.
- Liesenfeld R. [2001], A Generalized Bivariate Mixture Model for Stock Price Volatility and Training Volume, Journal of Econometrics, 11, s. 141-178.
- Maddala G.S. [1983], Limited-dependent and Qualitative Variables in Econometrics, Cambridge University Press, Cambridge.
- Magiera R. [2005], Modele i metody statystyki matematycznej. Część I: Rozkłady i symulacje stochastyczne, Oficyna wydawnicza GiS, Wrocław.

- Marin J.M., K. Mengersen, C.P. Robert [2005], Bayesian Modelling and Inference on Mixtures of Distributions, [w:] Handbook of Statistics (Bayesian Thinking, Modeling and Computation), 25, red. D. Dey, C.R. Rao), Elsevier-Sciences.
- Marzec J. [2006], Bayesowski model wielomianowy z rozkładem t Studenta dla kategorii uporządkowanych, Metody ilościowe w naukach ekonomicznych (red. A. Welfe), Wydawnictwo SGH w Warszawie, s. 123-144
- McFadden D.L. [1984], Econometrics Analysis of Qualitative Response Models, [w:] Handbooks of Econometrics vol II, red.Z. Griliches, M. Intriligator, Elsevier Science Publishers.
- Newton M.A., A.E. Raftery [1994], Approximate Bayesian Inference by the Weighted Likelihood Bootstrap (with discussion), Journal of the Royal Statistical Society B, 56, s. 3-48.
- O'Hagan A. [1994], Bayesian Inference, J. Wiley, New York.
- Osiewalski J., M.F.J. Steel [1993] A Bayesian Perspective on Model Selection (maszynopis); opublikowano w j. hiszpańskim: Una perspectiva bayesiana en selección de modelos, Cuadernos Economicos, 55/3.
- Osiewalski J. [1991] Ekonometria bayesowska w zastosowaniach, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.
- Osiewalski J., J. Marzec [2004], Model dwumianowy II rzędu i skośny rozkład Studenta w analizie ryzyka kredytowego, Folia Oeconomica Cracoviensia, vol. 45, s. 63-84
- Pearson K. [1894] Contributions to the Mathematical Theory of Evolution, Philosophical Transactions of the Royal Society of London A, 185, 71-110.
- Qu P., Y. Qu [2000], A Bayesian Approach to Finite Mixture models in Bioassay via Data Augmentation and Gibbs Sampling and Its Application to Insecticide Resistance, Biometrics, 56, s. 1249-1255.
- Quandt R.E., J.B. Ramsey [1978], Estimating Mixtures of Normal Distributions and Switching Regressions, Journal of the American Statistical Association, vol. 73, s. 730-738.
- Richardson S., P.J. Green [1997], On Bayesian Analysis of Mixture with an Unknown Number of Components, Journal of the Royal Statistical Society B, vol. 59, s. 731-792.
- Roeder K., L. Wasserman [1997], Practical Bayesian Density Estimation Using Mixtures of Normals, Journal of the American Statistical Association, vol. 92, nr. 439; s. 894-902.
- So M.K.P., K. Lan, W.K Li [1989], A Stochastic volatility model with Markov Switching, Journal of Business and Economic Statistics, nr 16, s. 244-253.
- Tierney L. [1994], Markov Chains for Exploring Posterior Distributions (with discussion), Annals of Statistics, 22, s. 1701-1762.
- Titterton D.M., A.F.M. Smith, U.E. Makov [1985], Statistical Analysis of Finite Mixture Distributions, John Wiley, New York.